

3 利用者コメントによるウェブサイトの人気度推定

会津大学

コンピュータ理工学部 情報システム学部門

准教授 成瀬 継太郎

■新技術の概要

任意のウェブサイトの人気度を、利用者コメントの人数と件数から推定する。本技術では人間の行動特性と統計的な解析により、従来把握が困難であったコメントを書かずに見るだけの層の人気度も推定可能となった。

■従来技術・競合技術との比較

ウェブサイトの人気度の代表的な調査方法は抽出調査であるが、調査コストが高く、抽出の偏りの影響も大きかった。本技術はウェブ上の誰もがアクセスできる情報を用いるため、安価で偏りが少ないという利点を持つ。

■新技術の特徴

- ・利用者コメントという「クチコミ」からインターネット視聴率を推定
- ・ブログ、SNS、掲示板といったネット上でのコメントから、ネット外の物事に興味を持つユーザー数を推定
- ・カスタマ・サービスにおいて、一部のカスタマからのクレームから潜在的な全体の不満率の推定

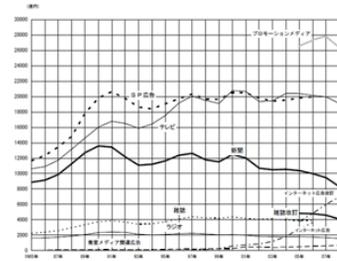
■想定される用途

- ・インターネット視聴率調査
- ・インターネット広告効果調査
- ・ウェブサーバの負荷予測

利用者コメントによる ウェブサイトの人気度推定

会津大学 コンピュータ理工学部
准教授 成瀬継太郎

研究背景



媒体別広告費の移り変わり, 電通, 2009年2月23日
<http://www.dentsu.co.jp/news/release/2009/pdf/2009013-0223.pdf>

研究背景

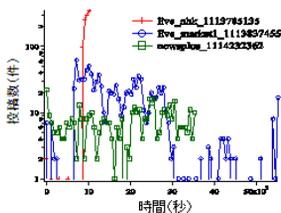
- インターネット広告市場の拡大
 - 2008年 5,373億円, 前年比116.3%, 全広告の10.4%
 - 今後も成長が見込まれる
- インターネットでの広告効果の調査と検証が必要
 - 適切な広告出稿費の算出
 - インターネット視聴率 = 各ページの閲覧数の調査

研究背景

- 各ページの閲覧者数の調査方法
 - (1) ウェブサーバ運用機関によるログの調査
 - (2) 専門の第三者機関によるサービス
 - ユーザの抽出検査
 - Alexa, ビデオリサーチ, ネットレイティングスなど
 - これらは一部の人のみだけが知れる情報
 - それらの情報を検証する第三の手法が必要

新技術の基となる研究成果・技術

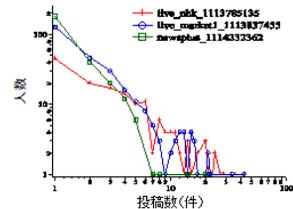
あるウェブ掲示板における異なる3つのスレッド(話題)の
単位時間当たりの投稿数の推移



共通の特徴は見つからない

新技術の基となる研究成果・技術

先ほどの3つのスレッドの投稿数とその人数の分布(投稿回数)



共通の特徴: 「ロングテール」

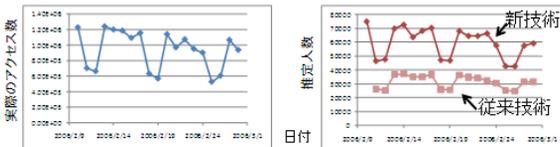
新技術の基となる研究成果・技術

- 研究成果
 - 投稿次数の分布:「メッセージを投稿した人」と
 - 閲覧者数:「メッセージを読むだけの人」の関係を
 - 数理的に導出
- それを応用すると、
 - ウェブサイトに公開されているメッセージ数から
 - それを読んでいる母集団数を推定可能

新技術の基となる研究成果・技術

- 適用事例1
 - 新技術の基本的な推定能力を検証
 - ウェブ掲示板の投稿次数から推定される母集団数と
 - 実際の掲示板サーバへのアクセス数を比較
 - 2006年2月の2週目から4週目の
 - 3,128,155メッセージ / 33,069スレッドで検証

新技術の基となる研究成果・技術



実際の掲示板サーバへのアクセス数

推定された集団数

ポイント: グラフの形状
新技術による推定人数は、実アクセス数の大小関係をよく再現

新技術	従来技術
0.953	0.943

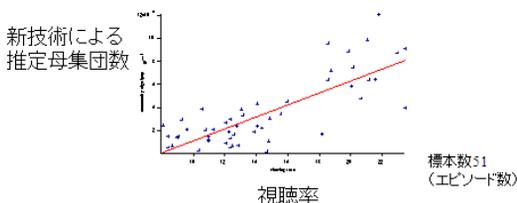
実アクセス数と推定集団数の相関係数
新技術により向上

新技術はより高い精度で推定可能

新技術の基となる研究成果・技術

- 適用事例2
 - 掲示板での人気度と他のメディアでの人気度の関連を調査
 - 推定される母集団数と番組視聴率を比較
 - 2006年の12月から2007年3月までのTVドラマと
 - 関連する94スレッド

新技術の基となる研究成果・技術



異なるメディアにも関わらず、
推定数と視聴率は高い相関 (0.799)

従来技術とその問題点

- ウェブサイトの人気度 (= ネット視聴率) の代表的な調査方法
 - 抽出調査
 - 調査コストが高い
 - 抽出の偏りの影響も大きい
 - ウェブサーバ運用機関によるログの調査
 - ユニーク・ユーザ数の推定が困難
 - 視聴者の属性が不明

新技術の特徴・従来技術との比較

- 新技術の利点
 - ウェブ上の誰もがアクセスできる情報を用いるため、安価で偏りが少ない
 - 異なるウェブサイトであっても比較可能
 - ページ単位での人気度の推定が可能
 - 異なるメディアに対しても、ウェブでの評判から人気度を推定可能

13

新技術の特徴・従来技術との比較

- 新技術の利点
 - 小規模なシステムで実現可能
 - 1台のPC(クライアント機能)とインターネット接続があれば実装可能
 - 対象を調査する規模や頻度に応じて、PCの台数や性能、データ保存容量を増強

14

想定される用途

- 誰もが利用可能な情報から任意のウェブサイトの人気度が推定できるため、**新しいインターネット視聴率調査サービスの提供**
 - 抽出調査ではなく、実際にWebサイトを閲覧している多人数のデータから推定するため、精度向上が期待

15

想定される用途

- **ターゲット1: 広告出稿者**
 - 広告効果の検証・調査
 - 広告費、広告媒体の調査・評価
- **ターゲット2: Webサイト運用者**
 - サーバの負荷分散へ応用
- **ターゲット3: 広告代理店**
 - 新しい広告効果の評価方法
 - 宣伝効果の透明化

16

想定される業界

- **利用者・対象者**
 - インターネット広告関連
- **市場規模**
 - 媒体費5,373億円, 製作費1,610億円

17

実用化に向けた課題

- 実用化に向けて、長時間運用可能なソフトウェアシステムの構築
- より広範なインターネットメディアでの推定能力の検証実験

18

企業への期待

- 実システムとして公開, 運用
 - 高信頼性のソフトウェアシステム, サービスの実装技術
 - 大量のカスタマ・データを用いた検証実験

19

本技術に関する知的財産権

- 発明の名称: 利用人数算出システム, 利用人数算出方法および利用人数算出プログラム
- 出願番号: 特願2006-337242
- 出願人: 公立大学法人会津大学
- 発明者: 成瀬継太郎
- (審査請求中)

20